

Modulus Computational Entropy

Maciej Skórski

February 11, 2013

Abstract

Recently, it has been shown by Pietrzak et al. that a leakage chain rule does not hold in general for commonly used definition of HILL Min-Entropy. We introduce the concept of *modulus computational entropy* and use it as a technical tool that allows to prove a chain rule for leakage. We show that the definition of modulus computational entropy is implied by several, sometimes seemingly unrelated assumptions, especially the ones already used in the literature in the context of leakage. Our results indicate that the concept of modulus entropy is, up to now, the weakest restriction that guarantees that the chain rule works.

1 Introduction

Entropy is the most fundamental concept in Information Theory. First introduced in this context by Shannon [Sha48], as a measure of the uncertainty associated with a probability distribution, it has been generalized in many ways. The commonly used generalization of Shannon Entropy is Rényi Entropy, defined for any arbitrary nonnegative order, which includes Shannon Entropy as a special case of order 1. Informally, a reasonable entropy measure indicates for a given distribution how much randomness it contains. According to this intuition, distributions uniform over large sets should have very high entropy, in opposite to distributions which has small support or hit a small set with high probability being easy to predict.

Indistinguishability and entropy. The notion of entropy has been generalized also for the purpose of Computational Complexity Theory and Cryptography, to take *computational* aspects into account. The reader might wish to refer to [Rey11] for a short survey. Historically computational entropy was first introduced in [Yao82] and, basing on a different concept, in [HILL99]. This last approach, based on the notion of indistinguishability, is the one we follow in this work. Let us try to give some intuitions here (the precisely definitions will be given in Section 2). To define computational entropy of X , one relaxes the requirement that X should have entropy itself. Instead, we assume that X is only close to a distribution Y which has suitable information-theoretic entropy. To make this work we have to specify two things: (a) the entropy we use and (b) what does it mean "being close". We note that due to technical reasons, special attention is drawn usually to the case of Rényi Entropy of order ∞ , called min-entropy. Min-entropy can be simply characterized by the unpredictability property, since it is nothing more than just the logarithm of the probability of most likely taken value, taken with a minus sign. To give a rigorous formulation of (b), one uses a concept of indistinguishability, being in fact the same concept as separation in Convex Analysis or Topology. Namely, we say that a function D separates (*distinguishes*) a set \mathbb{X} from another set \mathbb{Y} with *advantage* at least ϵ if $D(x) - D(y) \geq \epsilon$ for every $x \in \mathbb{X}$, $y \in \mathbb{Y}$. In turn, for a predefined class \mathcal{D} of functions, two sets are said to be (\mathcal{D}, ϵ) -indistinguishable, if there is no $D \in \mathcal{D}$ that can distinguish between these two sets with advantage greater than ϵ . The smaller ϵ and wider class \mathcal{D} we take, the stronger indistinguishability we obtain. Especially, indistinguishability applied to two probability distributions (as one-element sets) and the class of all boolean functions (as distinguishers), meaning acting D on a probability distribution \mathbf{P}_X as taking expected value $D(\mathbf{P}_X) = \mathbf{E}_{x \leftarrow X} D(x)$, yields the definition statistical distance. In applications involving computational complexity, one usually use circuits of bounded size as a distinguishers class.

Leakage Lemma and a Chain Rule. Leakage lemma is the term commonly used in referring to various generalizations of the observation which, saying less formally, states that min-entropy of a distribution X conditioned on another distribution Z distributed over $\{0, 1\}^m$ decreases, with respect to min-entropy of X , by at most m (the number of bits in the string encoding Z). The name comes from security-related applications, where one considers entropy of a distribution conditioned on information that might have been revealed to the adversary. The larger difference between entropy of a distribution and entropy of the corresponding conditioned distribution, the larger leakage is; such an approach, based on computational entropy, was used first by Dziembowski and Pietrzak in [DP08]. In turn, the term leakage chain rule is used to state the same principle for the case when we are given entropy of an already conditioned distribution and we are conditioning it on yet another distribution. Such further conditioning of an already conditioned distribution refers to so called "leakage-after-leakage" scenario.

Although the leakage chain rule is very easy to prove in the information-theoretic framework for conditional min-entropy or even smooth min-entropy (in fact also for Rényi entropy of an arbitrary order), the problem appears in the case of computational generalizations of entropy.

The computational leakage lemma [DP08, FR11], turned out not to give rise naturally to the leakage chain rule at least for important indistinguishability based definitions of conditional computational entropy and is addressed as an open problem [FOR12]. A computational leakage chain rule was proved only for specific scenarios, or by adding strong assumptions to definitions ([FR11], [CKLR11]) or by using slightly changed definitions (see [Rey11] for the discussion of computational relaxed entropy based on Leakage Lemma [GW10]). Recently, a counterexample has been shown ([SK12]) to the chain rule for computational min entropy.

Our contribution. Interested in establishing the possible weakest additional assumptions to make the leakage chain rule work for standard (defined via indistinguishability based on min-entropy) computational entropy, we define the modulus computational entropy and show that its definition is satisfied by technical requirements which have been used by other authors to prove a chain rule: the decomposable entropy introduced by Fuller and Reyzin [FR11] and the samplability assumption used by Kai-Min Chung et al. in [CKLR11]. Furthermore, we investigate three cases that has not been considered yet: (a) the case where computational entropy is sufficiently high, (b) the existence of an **NP** oracle to which distinguishers are given access, and (c) the case when the leakage is relatively short. In all these cases our definition is fulfilled and the chain rule works. Summing up, while we cannot solve this problem in general, we solve it for a few important concrete cases and reduce these, together with already known solutions, to the one single concept.

Outline of the work. Section 2 deals with some preliminary concepts, conventions and notations. In Section 3 we explain basic definitions and terminology being used in the case of computational entropy. In Section 3 we also discuss the cases where a chain rule is known to work. In Section 4 we define the modulus entropy and show that for modulus entropy the leakage chain rule holds. Section 5 contains a brief summary of the most important consequences of our results - partial solutions for the chain rule problem. Section 6 gives proofs of the conversion to modulus entropy for some cases, especially for the ones already having been considered in the literature in the context of leakage. Section 7 contains proofs of some technical results.

2 Preliminaries

Throughout this work we assume that all random variables are defined on some finite probability space and they take values in $\{0, 1\}^*$. If X is a random variable then \mathbf{P}_X will be its distribution. When the context is clear we will sometimes slightly abuse the notation and denote \mathbf{P}_X by X . Writing $X \in S$ we mean that X takes its values in the set S . By $|S|$ we denote the cardinality of S . For two random variables X, Z by $X|Z = z$ we denote the distribution of X conditioned on $Z = z$ and (X, Z) means the concatenation of X and Z . For every n , by U_n we denote the uniform distribution over $\{0, 1\}^n$. By $(\text{det}\{0, 1\}, s)$ and $(\text{det}[0, 1], s)$ we mean the class of all deterministic circuits of size at most s , with output in the set $\{0, 1\}$ and $[0, 1]$ respectively. Similarly, we denote by $(\text{rand}\{0, 1\}, s)$ the set of all randomized boolean circuits of size at most s . All algorithms are taken to the base 2. We say that function $\mu : \mathcal{X} \rightarrow \mathbb{R}$ is a convex combination of functions $\mu_i : \mathcal{X} \rightarrow \mathbb{R}$ if $\mu = \sum_{i=1}^l \alpha_i \mu_i$ for some nonnegative numbers α_i satisfying $\sum_i \alpha_i = 1$. For $D : \mathcal{X} \rightarrow [0, 1]$ and $k \leq \log |\mathcal{X}|$ we denote by $\text{Max}_D^k \subseteq \mathcal{X}$ a set of cardinality 2^k such that for every $x \in \text{Max}_D^k$ and every $x' \notin \text{Max}_D^k$ we have $D(x) \geq D(x')$. For $D : \mathcal{X} \rightarrow \{0, 1\}$ we define $|D| = \sum_{x \in \mathcal{X}} D(x)$.

2.1 Min Entropy

We start with recalling information-theoretic notions.

Definition 1 (Min Entropy). Given a random variable X we say that it has at least k bits of min-entropy and denote by $\mathbf{H}_\infty(X) \geq k$ if and only if $\max_x \mathbf{P}_X(x) \leq 2^{-k}$

The *conditional* min-entropy can be defined in two ways, both compatible with the above definition. The first one is given below.

Definition 2 (Worst-Case Conditional Min-Entropy). Given a pair of random variables (X, Z) we say that X *conditioned on Z has min-entropy at least k* and denote $\mathbf{H}_\infty(X|Z) \geq k$, if and only if $\max_x \mathbf{P}_{X|Z=z}(x) \leq 2^{-k}$ for every z

It is called the *worst-case* because it requires X to have high min-entropy when it is conditioned on an event “ $Z = z$ ” for *every* z . The alternative definition requires this fact to hold *on average*:

Definition 3 (Average Conditional Min-Entropy). Given a pair of random variables (X, Z) we say that X *conditioned on Z has average min-entropy at least k* and denote $\tilde{\mathbf{H}}_\infty(X|Z) \geq k$, if and only if $\mathbf{E}_{z \leftarrow Z} \left[\max_x \mathbf{P}_{X|Z=z}(x) \right] \leq 2^{-k}$

Usually it is not so important which of these definitions is used, because one can convert (via a Markov-type argument) the average conditional min entropy to the worst case variant.

Lemma 1 (See [DORS08], Lemma 2.2). *Suppose that $\tilde{\mathbf{H}}_\infty(X|Z) \geq k$. Then holds $\mathbf{H}_\infty(X|Z = z) \geq k - \log \frac{1}{\delta}$ with probability at least $1 - \delta$ over $z \leftarrow Z$.*

2.2 Indistinguishability

Below we outline the concept of *indistinguishability*, being a key point in defining computational entropy in the next section.

Definition 4. Let \mathbb{X} and \mathbb{Y} be subsets of some set \mathcal{P} . Given a positive real number ϵ we say that a function $F : \mathcal{P} \rightarrow [0, 1]$ *distinguishes between \mathbb{X} and \mathbb{Y} with advantage at least ϵ* if

$$\text{for every } x \in \mathbb{X} \text{ and } y \in \mathbb{Y} \text{ we have } |F(x) - F(y)| \geq \epsilon.$$

Definition 5. Let \mathbb{X} and \mathbb{Y} be as in Definition 4. Given a class \mathcal{F} consisting of $[0, 1]$ -valued functions on \mathcal{P} , we say that \mathbb{X} and \mathbb{Y} are (\mathcal{F}, ϵ) -*indistinguishable* if there is no $F \in \mathcal{F}$ that can distinguish between \mathbb{X} and \mathbb{Y} with advantage greater than ϵ .

In this paper we are mostly interested in a special case when \mathcal{P} is equal to the set of all probability distributions over some finite space Ω . In this case, every function $D : \Omega \rightarrow [0, 1]$ gives rise to a distinguisher $F_D : \mathcal{P} \rightarrow [0, 1]$ defined as:

$$F_D(\mu) = \mathbf{E}_{x \leftarrow \mu} D(x).$$

Thus, we will overload the notation and say that D *distinguishes between \mathbb{X} and \mathbb{Y} with advantage at least ϵ* if the corresponding function F_D distinguishes between \mathbb{X} and \mathbb{Y} with advantage at least ϵ . We note that D can also be a *randomized* function, which can be modeled by giving to D an additional input R chosen independently at random. In this case, the expected value in the definition above is taken also over the choice of R .

3 Computational Entropy and Leakage - previous works

As mentioned in the introduction, computational entropy can be obtained by generalizing min-entropy (or other notion of entropy) in many ways. We follow the approach based on indistinguishability as it seems to be the most standard way and was originally used for studying leakage [DP08] as well as further leakage-related results [CKLR11, FR11, GW10].

3.1 Defining Computational Entropy

Three-layer definition. There are three key points, essential for defining computational entropy via indistinguishability:

- (a) specify, for every k , what it means that a distribution “has (non-computational) entropy at least k ”,
- (b) model the adversary, in particular define his computational power, and determine his maximal acceptable success probability, and
- (c) define the measure of the “computational distance” between a given distribution and the set of distributions with entropy at least k (in the sense of (a)).

In (a) one usually uses information-theoretic notion of entropy, most often the min-entropy¹. For (b) one uses a pair (\mathcal{D}, ϵ) within the framework described in Section 2.2. Finally, a rigorous formulation of (c) can be given in two ways, traditionally called the “HILL” or the “Metric” versions. In the HILL version, while defining entropy of a random variable X , we require X to be indistinguishable from *some* distribution with high entropy (in the sense of (a)), whereas in the definition of the Metric Entropy we require X to be indistinguishable from the set of *all* of high-entropy distributions, which is a bit *weaker* assumption. The formal definitions below are provided for the conditional versions of both notions. The unconditional versions, denoted $\mathbf{H}^{\text{HILL}, \mathcal{D}, \epsilon}(X)$ and $\mathbf{H}^{\text{Metric}, \mathcal{D}, \epsilon}(X)$, are special cases of these notions obtained by fixing in the definitions below Z to be constant.

Definition 6 (HILL Computational Worst-Case Conditional Entropy). Let X, Z be random variables taking values in $\{0, 1\}^n$ and $\{0, 1\}^m$ respectively. Given $\epsilon > 0$, and a class of distinguishers \mathcal{D} , we say that X *conditioned on Z has at least k bits of computational HILL entropy against (\mathcal{D}, ϵ)* and denote $\mathbf{H}^{\text{HILL}, \mathcal{D}, \epsilon}(X|Z) \geq k$ if *there exists* a random variable $Y \in \{0, 1\}^n$ satisfying $\mathbf{H}_\infty(Y|Z) \geq k$, such that (X, Z) is (\mathcal{D}, ϵ) -indistinguishable from (Y, Z) .

Definition 7 (Metric Computational Worst-Case Conditional Entropy). With ϵ, \mathcal{D}, X and Z as in Def. 6, we say that X *conditioned on Z has at least k bits of computational metric entropy against (\mathcal{D}, ϵ)* and denote $\mathbf{H}^{\text{Metric}, \mathcal{D}, \epsilon}(X|Z) \geq k$ if (X, Z) is (\mathcal{D}, ϵ) -indistinguishable from the set of *all* distributions (Y, Z) , satisfying $\mathbf{H}_\infty(Y|Z) \geq k$.

The definitions of the HILL Computational *Average*-Case Conditional Entropy $\tilde{\mathbf{H}}^{\text{HILL}, \mathcal{D}, \epsilon}(X|Z)$ and the Metric Computational Worst-Case Conditional Entropy $\tilde{\mathbf{H}}^{\text{Metric}, \mathcal{D}, \epsilon}(X|Z)$ are obtained by replacing $\mathbf{H}_\infty(Y|Z) \geq k$ in Def. 6 and Def. 7 (resp.) with $\tilde{\mathbf{H}}_\infty(Y|Z) \geq k$. We note that one usually uses a different formulation of the definitions of Metric and HILL Entropy.

Definition 6: $\mathbf{H}^{\text{HILL}, \mathcal{D}, \epsilon}(X|Z) \geq k$ if there exists a random variable $Y \in \{0, 1\}^n$ such that for every $D \in \mathcal{D}$, we have $|\mathbf{E}_{(x,z) \leftarrow (X,Z)} D(x, z) - \mathbf{E}_{(x,z) \leftarrow (Y,Z)} D(x, z)| \leq \epsilon$.

Definition 7: $\mathbf{H}^{\text{Metric}, \mathcal{D}, \epsilon}(X|Z) \geq k$ if for every $D \in \mathcal{D}$ there exists a random variable $Y \in \{0, 1\}^n$ such that $|\mathbf{E}_{(x,z) \leftarrow (X,Z)} D(x, z) - \mathbf{E}_{(x,z) \leftarrow (Y,Z)} D(x, z)| \leq \epsilon$.

It is not hard to verify that both formulations are equivalent. However, our, more general, definitional approach appears to be more useful for the applications presented in the sequel.

The equivalence between HILL and Metric-type Entropy. The Metric entropy, which was introduced after the HILL one, is more convenient for proving leakage-related results. Both versions are very close in practical applications as there exists a conversion from Metric Entropy (against real valued circuits) to HILL entropy [BSW03]. This result in its full generality can be stated as follows

¹We use only min-entropy in this work. See, however, [VZ12] for a similar definitions based on Shannon Entropy.

Theorem 1 (Generalization of [BSW03], Thm. 5.2). *Let \mathcal{P} be the set of all probability measures over Ω . Suppose that we are given a class \mathcal{D} of $[0, 1]$ -valued functions on Ω , with the following property: if $D \in \mathcal{D}$ then $D^c \stackrel{\text{def}}{=} \mathbf{1} - D \in \mathcal{D}$. For $\delta > 0$, let \mathcal{D}' be a class consisting of all convex combinations of length $\mathcal{O}\left(\frac{\log |\Omega|}{\delta^2}\right)$ over \mathcal{D} . Let $\mathcal{C} \subset \mathcal{P}$ be any arbitrary convex subset of probability measures and $X \in \mathcal{P}$ be a fixed distribution. Consider the following statements:*

- (a) X is $(\mathcal{D}, \epsilon + \delta)$ indistinguishable from some distribution $Y \in \mathcal{C}$
- (b) X is (\mathcal{D}', ϵ) indistinguishable from the set of all distribution $Y \in \mathcal{C}$

Then (b) implies (a).

Remark 1. This result was formulated in [BSW03] in a less general form, namely $\Omega = \{0, 1\}^n$, \mathcal{C} is the set of distributions with min-entropy at least k , and $\mathcal{D}, \mathcal{D}'$ are the classes of $[0, 1]$ -valued circuits of size s and $\mathcal{O}\left(s \cdot \frac{n}{\delta^2}\right)$ respectively. The inspection of the proof shows that: (a) the chosen space Ω can be an arbitrary finite set, and the number n appearing in the assertion is equal to $\log |\Omega|$, (b) the chosen set \mathcal{C} can be replaced by an arbitrary convex set of distributions, (c) the complexity of the class \mathcal{D}' is chosen only to ensure that \mathcal{D}' contains all convex combinations of length $\mathcal{O}\left(\frac{\log |\Omega|}{\delta^2}\right)$ of elements of \mathcal{C} .

Remark 2. By choosing $\Omega = \{0, 1\}^{n+m}$, a random variable $Z \in \{0, 1\}^m$ and \mathcal{C} to be the set of all distributions (Y, Z) satisfying $(Y, Z) : \mathbf{H}_\infty(Y|Z) \geq k$ or alternatively $\tilde{\mathbf{H}}_\infty(Y|Z) \geq k$, we obtain the conversion from Metric Conditional Entropy to HILL Conditional Entropy, for both: worst case and average case variants.

3.2 Leakage Rules

We are now ready to state the leakage chain rule for conditional min-entropy and compare it with its known generalizations to computational case. Generally, we are interested in the following problem:

Suppose we have a pair of random variables (X, Z_1) and we know the conditional entropy of X given Z_1 . What is the lower bound on the entropy of X given (Z_1, Z_2) , where Z_2 is some other (possibly correlated) random variable?

In the information-theoretic case we have the following estimate (cf. [DORS08], Lemma 2.2).

Lemma 2 (Leakage Chain Rule). *Let X, Z_1, Z_2 be random variables over $\{0, 1\}^n, \{0, 1\}^{m_1}, \{0, 1\}^{m_2}$ respectively. Then*

$$\tilde{\mathbf{H}}_\infty(X|Z_1, Z_2) \geq \tilde{\mathbf{H}}_\infty(X|Z_1) - m_2 \quad (1)$$

The name ‘‘Leakage Chain Rule’’ comes from the fact that we think of Z_1 and Z_2 as information about X that ‘‘leaked’’ subsequently to the adversary. In the computational framework, the first leakage-related result appeared in [DP08] and was improved next in [FR11]. It is called Leakage Lemma as it deals with the case of one leakage only.

Lemma 3 (Leakage Lemma [FR11]). *Let X and Z be random variables over $\{0, 1\}^n$ and $\{0, 1\}^m$, resp. Then*

$$\tilde{\mathbf{H}}^{\text{Metric}, [0, 1], s', \epsilon'}(X|Z = z) \geq \mathbf{H}^{\text{Metric}, [0, 1], s, \epsilon}(X) - m$$

where $s' = s + \mathcal{O}(1)$ and $\epsilon' = 2^m \epsilon$.

Let us observe, at least under assumption that there exists an exponentially secure pseudorandom generator, that both losses: in quantity (by m bits) and security measured as s/ϵ (by factor almost equal to 2^m) can appear simultaneously²; see Theorem 9 in this work.

²In [FR11] the authors leave this problem as an open question

It is a natural question to ask if the Leakage Chain Rule (Lemma 2) can be “translated” into the computational version. In particular, one might be tempted to conjecture that for X, Z_1 and Z_2 as in Lemma 2 it holds that

$$\tilde{\mathbf{H}}^{\text{Metric}, [0,1], s', \epsilon'}(X|Z_1, Z_2) \geq? \tilde{\mathbf{H}}^{\text{Metric}, [0,1], s, \epsilon}(X|Z_1) - m_2, \quad (2)$$

with security loss of factor 2^{m_2} , where by security loss we mean $\frac{s'}{\epsilon'}/\frac{s}{\epsilon}$ (which reduces to ϵ/ϵ' if $s' \approx s$). Unfortunately, this conjecture is still unproven in its full generality [FOR12]. On the positive side, some progress towards proving it has been recently made in [FR11] and [CKLR11] where it is proven for restricted classes of entropies. In [FR11] this restriction is called *decomposability*. More precisely, their definition is as follows.

Definition 8 ([FR11]). Let X, Z be as in Lemma 3. We say that X has *decomposable metric-entropy conditioned on Z at least k* and denote by $\mathbf{H}^{\text{Metric-d}, [0,1], s, \epsilon}(X|Z) \geq k$, if for every z

$$\mathbf{H}^{\text{Metric}, [0,1], s, \epsilon(z)}(X|Z = z) \geq k(z)$$

where $\epsilon(z)$ and $k(z)$ are numbers satisfying $\mathbf{E}_{z \leftarrow Z} 2^{-k(z)} = 2^{-k}$ and $\mathbf{E}_{z \leftarrow Z} |\epsilon(z)| \leq \epsilon$.

Using this definition they are able to prove the following.

Theorem 2 ([FR11]). Let X, Z_1, Z_2 be as in Lemma 2. Then for $s' \approx s$, and $\epsilon' = 2^{m_2}\epsilon$, we have

$$\tilde{\mathbf{H}}^{\text{Metric-d}, [0,1], s', \epsilon'}(X|Z_1, Z_2) \geq \tilde{\mathbf{H}}^{\text{Metric-d}, [0,1], s, \epsilon}(X|Z_1) - m_2$$

In the other approach [CKLR11], the authors assume the existence of an effectively samplable distribution with high conditional min-entropy being indistinguishable from (X, Z_1) . The precise formulation of their result is given below

Theorem 3 ([CKLR11]). Let X, Z_1, Z_2 be as above. Suppose that there exists a random variable Y' with the following properties: (a) $\mathbf{H}_\infty(Y'|Z_1) \geq k$ and (b) there exists a randomized circuit Γ receiving on its input $z \in \text{supp}(Z_1)$ and returning samples from $Y'|Z_1 = z_1$. Then

$$\mathbf{H}^{\text{Metric}, [0,1], s', \epsilon'}(X|Z_1, Z_2) \geq \mathbf{H}^{\text{Metric}, [0,1], s, \epsilon}(X|Z_1) - |Z_2| - \log \frac{1}{\delta},$$

for $s' = \Omega\left(s \cdot \frac{\delta}{2^{m_2}} - s_O\right)$, $\epsilon' \approx \epsilon + \delta$.

We note that there is yet another result related to the chain rule problem, due to [GW10]. The authors prove a version of 3 for a slightly different definition of Metric Conditional Min-Entropy. The difference is in Layer (a) of the definition: they require (X, Z) , to be indistinguishable from all distribution (Y, Z') satisfying $\mathbf{H}_\infty(Y|Z') \geq k$, where— in comparison to Definition 6— Z' is *not necessarily* equal to Z . As observed in [Rey11], one can easily generalize their approach to prove an “efficient” computational version of 2 for this definition, with a loss of a factor at most $\text{poly}(2^{m_2}, \epsilon^{-1})$ in security. It seems however, that in the context of leakage Definition 7 is more suitable [CKLR11].

4 Modulus Entropy

Our definition of modulus entropy is a bit different than Definition 8.

Definition 9 (Modulus Metric Entropy). Let $X \in \{0, 1\}^n$ and $X \in \{0, 1\}^m$ be random variables. Given $\epsilon > 0$ and a class of deterministic boolean functions \mathcal{D} , we say that X *conditioned on Z has modulus entropy at least k against (\mathcal{D}, ϵ)* , and denote it by $\mathbf{H}^{|\text{Metric}|, \mathcal{D}, \epsilon}(X|Z) \geq k$, if for any $D \in \mathcal{D}$ there exists a random variable $Y \in \{0, 1\}^n$, satisfying $\tilde{\mathbf{H}}_\infty(Y|Z) \geq k$, such that

$$\mathbf{E}_{z \leftarrow Z} |\mathbf{E}_{x \leftarrow (X|Z=z)} D(x, z) - \mathbf{E}_{x \leftarrow (Y|Z=z)} D(x, z)| \leq \epsilon \quad (3)$$

We emphasize that the above definition, being formulated for the worst-case conditional entropy, can be stated also for the average version which is obtained just by replacing \mathbf{H}_∞ with $\tilde{\mathbf{H}}_\infty$. By use of Lemma 1 we obtain immediately a conversion (with some loss) between them:

Lemma 4. *Suppose that $\tilde{\mathbf{H}}^{[\text{Metric}], \mathcal{D}, \epsilon}(X|Z) \geq k$. Then $\mathbf{H}^{[\text{Metric}], \mathcal{D}, \epsilon+\delta}(X|Z) \geq k - \log \frac{1}{\delta}$*

Some intuitions behind modulus entropy. The only difference between Definition 7 and Definition 9 is that they differ *in order of expectation and absolute value signs*. Thus, by the triangle inequality, the Modulus Entropy is smaller than Metric Entropy. However, they are not necessarily equal in general. Indeed, for D distinguishing between (X, Z) and (Y, Z) with the advantage no greater than ϵ , contributions to this advantage from particular values of z , given by the expressions $\epsilon_D(z) = \mathbf{E}_{x \leftarrow X|Z=z} D(x, z) - \mathbf{E}_{x \leftarrow Y|Z=z} D(x, z)$ can differ in signs. Hence, although we have $|\mathbf{E}_{z \leftarrow Z} \epsilon_D(z)| \leq \epsilon$, it does not imply $\mathbf{E}_{z \leftarrow Z} |\epsilon_D(z)| \leq \epsilon$ required by inequality (3). In comparison to Definition 8, our approach is far more general as allow numbers $\epsilon(z)$ as well as $k(z)$ (in the average variant) to be dependent on a chosen D . From a technical point of view, both definitions are formulated to control contributions from particular outcomes of Z to the parameter ϵ .

4.1 Leakage Chain Rule for Modulus Entropy

We now show how modulus entropy allows us to prove a leakage chain rule. We start with the reformulation of the leakage lemma proved in [FR11].

Lemma 5 (Corollary from [FR11]). *Let D be a boolean function and (X, Z) be as in Thm. 3. Suppose that $|\mathbf{E}_{x \leftarrow X} D(x) - \mathbf{E}_{x \leftarrow Y} D(x)| \leq \epsilon$, where $\mathbf{H}_\infty(Y) \geq k$. Then for any $z \in \text{supp}(Z)$ there exist a distribution Y'_z with min-entropy at least $k(z) = k - \log \frac{1}{\mathbf{P}(Z=z)}$ such that $|\mathbf{E}_{x \leftarrow X|Z=z} D(x) - \mathbf{E}_{x \leftarrow Y'_z} D(x)| \leq \frac{\epsilon}{\mathbf{P}(Z=z)}$.*

Now we are in position to prove the following chain rule, achieving the optimal parameters.

Theorem 4. *Let X, Z_1, Z_2 be as in Thm. 2 and \mathcal{D} be a class of boolean functions. Suppose that $\tilde{\mathbf{H}}^{[\text{Metric}], \mathcal{D}, \epsilon}(X|Z_1) \geq k$. Then $\tilde{\mathbf{H}}^{[\text{Metric}], \mathcal{D}, 2^{m_2}\epsilon}(X|Z_1, Z_2) \geq k - m_2$.*

Proof. Fix a distinguisher $D = D(x, z_1, z_2)$. We will construct a distribution (Y, Z_1, Z_2) such that $\tilde{\mathbf{H}}_\infty(Y|Z_1, Z_2) \geq k - m_2$ and D cannot distinguish (X, Z_1, Z_2) from (Y, Z_1, Z_2) with advantage better than $2^{m_2}\epsilon$.

For any z_2 , let (Y^{z_2}, Z_1) be a distribution corresponding to $D(\cdot, z_2)$, which existence is guaranteed by Definition 9 (we use notation Y^{z_2} to emphasize that this distribution depends also on z_2). More precisely, (Y^{z_2}, Z_1) is such that

$$\mathbf{E}_{z_1 \leftarrow Z_1} \underbrace{|\mathbf{E}_{x \leftarrow (X|Z_1=z_1)} D(x, z_1, z_2) - \mathbf{E}_{x \leftarrow (Y^{z_2}|Z_1=z_1)} D(x, z_1, z_2)|}_{\epsilon_D(z_1, z_2) :=} \leq \epsilon \quad (4)$$

holds (cf. (3) in Definition 9). For every pair (z_1, z_2) let $\epsilon_D(z_1, z_2)$ denote the value within the first expected value sign, as indicated on (4). Now, Lemma 5 implies that for any z_1, z_2 there exists a distribution Y'_{z_1, z_2} such that

$$|\mathbf{E}_{x \leftarrow X|(Z_1=z_1, Z_2=z_2)} D(x, z_1, z_2) - \mathbf{E}_{x \leftarrow Y'_{z_1, z_2}} D(x, z_1, z_2)| \leq \frac{\epsilon_D(z_1, z_2)}{\mathbf{P}(Z_2 = z_2|Z_1 = z_1)} \quad (5)$$

and its min-entropy $\mathbf{H}_\infty(Y'_{z_1, z_2})$ is at least $k(z_1, z_2)$, where

$$k(z_1, z_2) \geq \mathbf{H}_\infty(Y^{z_2}|Z_1 = z_1) - \log \frac{1}{\mathbf{P}(Z_2 = z_2|Z_1 = z_1)} \quad (6)$$

Let (Y, Z_1, Z_2) be a distribution given by $(Y | Z_1 = z_1, Z_2 = z_2) \stackrel{d}{=} Y'_{z_1, z_2}$. We now have

$$\begin{aligned}
& \leq \frac{\epsilon_D(z_1, z_2)}{\mathbf{P}(Z_2 = z_2 | Z_1 = z_1)} \quad (\text{by (5)}) \\
& \mathbf{E}_{(z_1, z_2) \leftarrow (Z_1, Z_2)} \left[\overbrace{\mathbf{E}_{x \leftarrow X | Z_1 = z_1, Z_2 = z_2} D(x, z_1, z_2) - \mathbf{E}_{x \leftarrow Y'_{z_1, z_2}} D(x, z_1, z_2)} \right] \\
& \leq \sum_{z_1, z_2} \mathbf{P}((Z_1, Z_2) = (z_1, z_2)) \cdot \frac{\epsilon_D(z_1, z_2)}{\mathbf{P}(Z_2 = z_2 | Z_1 = z_1)} \\
& = \sum_{z_1, z_2} \mathbf{P}(Z_1 = z_1) \epsilon_D(z_1, z_2) \\
& = \sum_{z_2} \mathbf{E}_{z_1 \leftarrow Z_1} \epsilon_D(z_1, z_2) \\
& \leq \sum_{z_2} \epsilon = 2^{m_2} \epsilon,
\end{aligned}$$

where the last inequality follows from (4). It remains to prove that $\tilde{\mathbf{H}}_\infty(Y | Z_1, Z_2) \geq k - m_2$. We have:

$$\begin{aligned}
\mathbf{E}_{(z_1, z_2) \leftarrow (Z_1, Z_2)} 2^{-k(z_1, z_2)} & \leq \mathbf{E}_{(z_1, z_2) \leftarrow (Z_1, Z_2)} \left[\max_x \mathbf{P}[Y^{z_2} = x | Z_1 = z_1] \cdot \frac{1}{\mathbf{P}[Z_2 = z_2 | Z_1 = z_1]} \right] \\
& = \sum_{z_1, z_2} \max_x \mathbf{P}[Y^{z_2} = x | Z_1 = z_1] \cdot \mathbf{P}[Z_1 = z_1] \\
& = \sum_{z_2} \mathbf{E}_{z_1 \leftarrow Z_1} \left[\max_x \mathbf{P}[Y^{z_2} = x | Z_1 = z_1] \right] \\
& \leq 2^{m_2} \cdot 2^{-k}
\end{aligned}$$

where the first step follows from (6) and the last one from $\tilde{\mathbf{H}}_\infty(Y^{z_2} | Z_1) \geq k$. \square

Chain Rule for entropy against different circuits classes Theorem 4 deals only with entropy against boolean deterministic distinguishers \mathcal{D} . It is natural to ask if one could replace this class with a more general one, in particular, would the theorem still hold if \mathcal{D} in its statement is equal to the class of randomized or real-valued distinguishers. We answer this question affirmatively in Lemma 6 below. To make its statement as strong as possible in its assumption we use the weakest possible option, which is the modulus entropy against boolean deterministic circuits, and in its assertion we use the strongest option i.e. the HILL entropy.³

Lemma 6. *Let X, Z be as in Theorem 3. Suppose that $\tilde{\mathbf{H}}^{\text{Metric}, s, \epsilon}(X | Z) \geq k$. In this case we have that $\mathbf{H}^{\text{HILL}, s', \epsilon'}(X | Z) \geq k'$, where $\epsilon' = \epsilon + 2\delta$, $s' = s \cdot \mathcal{O}\left(\frac{\delta^2}{n+m}\right)$ and $k' = k - \log \frac{1}{\delta}$.*

Proof of Lemma 6. If $\tilde{\mathbf{H}}^{\text{Metric}, s, \epsilon}(X | Z) \geq k$ then, as we pointed out in the discussion after Lemma 4, we have that $\tilde{\mathbf{H}}^{\text{Metric}, \det\{0,1\}, s, \epsilon}(X | Z) \geq k$. From Lemma 4 we obtain that $\mathbf{H}^{\text{Metric}, \det\{0,1\}, s, \epsilon + \delta}(X | Z) \geq k - \log \frac{1}{\delta}$. Applying Theorem 10, we obtain $\tilde{\mathbf{H}}^{\text{Metric}, s', [0,1], \epsilon + \delta}(X | Z) \geq k - \log \frac{1}{\delta}$ where $s' = s + \mathcal{O}(1)$. The claim follows now from Theorem 1 \square

Therefore, there is no meaningful loss in passing from Modulus Entropy to Metric Entropy, or even HILL Entropy. In the next section we consider some particular cases, where a conversion in the other direction is possible, up to negligible (in view of applications) loss.

³Recall that for the HILL Entropy all kinds of circuits: deterministic boolean, deterministic real valued, randomized boolean are equivalent [FR11] thus we can abbreviate the notation writing just $\mathbf{H}^{\text{HILL}, s', \epsilon'}(X | Z)$.

5 Passing to Modulus Entropy

While modulus entropy, as shown in Theorem 4, solves the leakage chain rule problem, it keeps being rather a technical assumption, nonequivalent to commonly used definitions. We will give some concrete examples where its definition is fulfilled, and thus admitting the chain rule. All of these examples, in comparison to the assertion of Theorem 4, rely on some another assumption added to metric entropy of $X|Z$. Conversion to the modulus entropy, meaning estimating the loss in parameters, is summarized in the table below.

Additional assumptions on $\tilde{\mathbf{H}}^{\text{Metric}, \{0,1\}, s, \epsilon}(X Z) \geq k$	Our conversion: $\tilde{\mathbf{H}}^{[\text{Metric}], s', \epsilon'}(X Z) \geq k'$			
	k'	ϵ'	s'	
(a) Decomposable entropy [FR11]	k	ϵ	s	Thm. 5
(b) Samplability of $Y Z = z$ given z , where $(Y, Z) \sim^{\epsilon, s}(X, Z)$ [CKLR11]	$k - \mathcal{O}(\log \frac{1}{\epsilon})$	$\mathcal{O}(\epsilon)$	$\mathcal{O}\left(\frac{s}{\epsilon^2}\right)$	Thm. 7
(c) Entropy against poly(n)-circuits, given an access to an NP oracle	$k - \mathcal{O}(\log \frac{1}{\epsilon})$	$\mathcal{O}(\epsilon^{\frac{1}{2}})$	$\mathcal{O}\left(\frac{s}{\text{poly}(n, \frac{1}{\epsilon})}\right)$	Thm. 8 (point b)
(d) Entropy very high, i.e. $k > n - \mathcal{O}(\log \frac{1}{\epsilon})$	$k - \mathcal{O}(\log \frac{1}{\epsilon})$	$\mathcal{O}(\epsilon^{\frac{1}{2}})$	$\mathcal{O}\left(\frac{s}{\frac{m+n}{\epsilon^3} \log \frac{1}{\epsilon}}\right)$	Thm. 8 (point a)
(e) None	k	$2^t \epsilon$	$s - \mathcal{O}(2^{m-t} m)$	Thm. 6

Table 1: Conversions to modulus entropy

As shown in the table, some of these assumptions have been already introduced in the literature to prove leakage-related results. The proofs of conversions will be given in the next section.

6 Proofs of conversion results

Throughout all the proofs in this section, X, Z are assumed to be random variables over $\{0, 1\}^n, \{0, 1\}^m$ respectively. The proofs are based on the following technical lemma.

Lemma 7. *Let X, Z be arbitrary random variables over $\{0, 1\}^n, \{0, 1\}^m$. Suppose that D is such that for all distributions (Y, Z) with $\mathbf{H}_\infty(Y|Z) \geq k$ the following holds:*

$$\mathbf{E}_{z \leftarrow Z} |\mathbf{E}_{x \leftarrow X|Z=z} D(x, z) - \mathbf{E}_{x \leftarrow Y|Z=z} D(x, z)| \geq \epsilon. \quad (7)$$

Then either for $D' = D$ or for $D' = D^c$ we have that for all distributions (Y, Z) with $\mathbf{H}_\infty(Y|Z) \geq k$ the following is true:

$$\mathbf{P}_{(x,z) \leftarrow (X,Z)} \left[D'(x, z) - \mathbf{E}_{x \leftarrow Y|Z=z} D'(x, z) \geq \frac{\epsilon}{4} \right] \geq \frac{\epsilon^2}{16}.$$

Proof. Consider the distribution (Y^+, Z) which minimizes the left-hand side of (7). Define

$$\epsilon(z) := |\mathbf{E}_{x \leftarrow X|Z=z} D(x, z) - \mathbf{E}_{x \leftarrow Y^+|Z=z} D(x, z)|.$$

Observe that

$$\begin{aligned} \min_{(Y,Z): \mathbf{H}_\infty(Y|Z) \geq k} \mathbf{E}_{z \leftarrow Z} |\mathbf{E}_{x \leftarrow X|Z=z} D(x, z) - \mathbf{E}_{x \leftarrow Y|Z=z} D(x, z)| &= \\ &= \mathbf{E}_{z \leftarrow Z} \left[\min_{Y_z: \mathbf{H}_\infty(Y_z) \geq k} |\mathbf{E}_{x \leftarrow X|Z=z} D(x, z) - \mathbf{E}_{x \leftarrow Y_z} D(x, z)| \right]. \end{aligned}$$

Therefore, for every distribution Y_z with min-entropy $\mathbf{H}_\infty(Y_z) \geq k$ we have

$$|\mathbf{E}_{x \leftarrow X|Z=z} D(x, z) - \mathbf{E}_{x \leftarrow Y|Z=z} D(x, z)| \geq \epsilon(z)$$

Note that if $\epsilon(z) > 0$ then either (a) $\mathbf{E}_{x \leftarrow X|Z=z} D(x, z) - \mathbf{E}_{x \leftarrow Y|Z=z} D(x, z) \geq \epsilon(z)$ or (b) $\mathbf{E}_{x \leftarrow X|Z=z} D(x, z) - \mathbf{E}_{x \leftarrow Y|Z=z} D(x, z) \leq -\epsilon(z)$ holds for all Y_z with $\mathbf{H}_\infty(Y_z) \geq k$. This follows from the convexity of the set of distributions $\mathbf{H}_\infty(Y_z) \geq k$, which in turn implies that all values of the expression $\mathbf{E}_{x \leftarrow X|Z=z} D(x, z) - \mathbf{E}_{x \leftarrow Y_z} D(x, z)$, over the choice of Y_z , form a convex set. Therefore

$$\mathbf{E}_{x \leftarrow X|Z=z} D'(x, z) - \mathbf{E}_{x \leftarrow Y_z} D'(x, z) \geq \epsilon(z)$$

holds for all Y_z with $\mathbf{H}_\infty(Y_z) \geq k$, where D' is defined, depending on z , by

$$D'(x, z) := \begin{cases} D(x, z) & \text{in case (a)} \\ D^c(x, z) & \text{in case (b)} \\ 0 & \text{if } \epsilon(z) = 0. \end{cases} \quad (8)$$

Since $\epsilon(z) \geq \frac{\epsilon}{2}$ holds⁴ with probability at least $\frac{\epsilon}{2}$ over $z \leftarrow Z$, we get

$$\mathbf{E}_{x \leftarrow X|Z=z} D'(x, z) - \max_{Y_z: \mathbf{H}_\infty(Y_z) \geq k} \mathbf{E}_{x \leftarrow Y|Z=z} D'(x, z) \geq \frac{\epsilon}{2}$$

with probability at least $\frac{\epsilon}{2}$ over $z \leftarrow Z$. For every such z we obtain

$$\mathbf{P}_{x \leftarrow X|Z=z} \left[D'(x, z) - \max_{Y_z: \mathbf{H}_\infty(Y_z) \geq k} \mathbf{E}_{x \leftarrow Y|Z=z} D'(x, z) \geq \frac{\epsilon}{4} \right] \geq \frac{\epsilon}{4}$$

Taking expectation over $z \leftarrow Z$ we conclude that

$$\mathbf{P}_{(x,z) \leftarrow (X,Z)} \left[D'(x, z) - \max_{Y_z: \mathbf{H}_\infty(Y_z) \geq k} \mathbf{E} D'(Y_z, z) \geq \frac{\epsilon}{4} \right] \geq \frac{\epsilon^2}{8}.$$

Therefore, for either $D' = D$, or for $D' = D^c$ the probability on the left-hand side of the above inequality needs to be at least $\frac{1}{2} \cdot \frac{\epsilon^2}{8} = \frac{\epsilon^2}{16}$, which proves the claim. \square

6.1 Decomposable entropy

We start with the trivial observation that Definition 8 is stronger than our Definition 9.

Theorem 5. *Suppose that $\tilde{\mathbf{H}}^{\text{Metric-d}, s, \epsilon}(X|Z) \geq k$. Then $\tilde{\mathbf{H}}^{\text{Metric}, s, \epsilon}(X|Z) \geq k$.*

Proof. Fix a distinguisher $D = D(x, z)$. According to Definition 8, for every z we have a distribution Y_z with min-entropy at least $k(z)$ such that $|\mathbf{E}_{x \leftarrow X_z} D(x) - \mathbf{E}_{x \leftarrow Y_z} D(x)| \leq \epsilon(z)$. Consider a distribution (Y, Z) defined by $(Y|Z = z) \stackrel{d}{=} Y_z$. Since $\mathbf{E}_{z \leftarrow Z} \epsilon(z) \leq \epsilon$, we obtain inequality (3). In turn, the assumptions on $k(z)$ implies $\tilde{\mathbf{H}}_\infty(Y|Z) \geq k$. \square

The following theorem converts Metric Entropy into Modulus Entropy (cf. case (e) in Table 5). Its principal significance is that the equivalence between both definitions is established, provided that Z is sufficiently short (grows at most logarithmically in the security parameters).

Theorem 6. *Suppose that $\mathbf{H}^{\text{Metric}, \{0,1\}, s, \epsilon}(X|Z) \geq k$. Then $\mathbf{H}^{\text{Metric}, s', \epsilon'}(X|Z) \geq k$, where $\epsilon' = 2^t \epsilon$ and $s' = s - \mathcal{O}(2^{m-t} m)$.*

Proof. For the sake of contradiction suppose that for some D of complexity s' and for every (Y, Z) such that $\mathbf{H}_\infty(Y|Z) \geq k$ we have that

$$\mathbf{E}_{z \leftarrow Z} |\mathbf{E}_{x \leftarrow X|Z=z} D(x, z) - \mathbf{E}_{x \leftarrow Y|Z=z} D(x, z)| \geq \epsilon'.$$

⁴Throughout the proofs, we will make use of the simple Markov-style principle: let X be a non-negative random variable bounded by M . Then $X > \frac{1}{2M} \mathbf{E}X$ with probability at least $\frac{1}{2} \mathbf{E}X$.

Applying the same reasoning as at the beginning of the proof of Lemma 7, we obtain that there exist a distinguisher D' (cf. (8)) such that for every Y_z with $\mathbf{H}_\infty(Y_z) \geq k$ it holds that

$$\mathbf{E}_{x \leftarrow X|z=z} D'(x, z) - \mathbf{E}_{x \leftarrow Y_z} D'(x, z) \geq \epsilon'(z), \quad (9)$$

where $\mathbf{E}_{z \leftarrow Z} \epsilon'(z) \geq \epsilon'$. Thus, for every distribution (Y, Z) with entropy $\mathbf{H}_\infty(Y|Z) \geq k$ we have

$$\mathbf{E}_{(x,z) \leftarrow (X,Z)} D'(x, z) - \mathbf{E}_{(x,z) \leftarrow (Y,Z)} D'(x, z) \geq \mathbf{E}_{z \leftarrow Z} \epsilon'(z) \geq \epsilon'.$$

Recall that in the proof of Thm. 7, the value $D'(x, z)$ is defined as equal to $D(x, z)$ or $D^c(x, z)$ or 0, depending on z . Instead, we can follow that construction with respect to only 2^{m-t} “heaviest” values z maximizing $\mathbf{P}(Z = z)\epsilon'(z)$ and setting $D' = 0$ for other z . The obtained circuit is of size at most $s' + \mathcal{O}(2^{m-t}m) = s$ and distinguishes with the advantage at least $2^{-t}\epsilon' = \epsilon$. \square

6.2 The samplability assumption

In the next theorem we deal with the samplability assumption used in [CKLR11].

Theorem 7. *Suppose that (X, Z) is (s, ϵ) -indistinguishable from a distribution (Y', Z) , with the following properties (a) $\mathbf{H}_\infty(Y'|Z) \geq k$ and (b) there exists a randomized circuit Γ receiving on its input $z \in \text{supp}(Z)$ and returning samples from the distribution $Y'|Z = z$. Then*

$$\mathbf{H}^{|\text{Metric}|, s \cdot \frac{\epsilon^2}{64} - \text{size}(\Gamma), 8\sqrt{\epsilon}}(X|Z) \geq k - 2 \log \left(\frac{1}{\epsilon} \right) - 7.$$

Proof. Suppose that $\mathbf{H}^{|\text{Metric}|, s', \epsilon'}(X|Z) < k'$, where $k' = k - 2 \log \left(\frac{1}{\epsilon} \right) - 7$ and $\epsilon' = \frac{\epsilon^2}{64}$ and $s' = \frac{\epsilon^2 s}{64} - \text{size}(\Gamma)$. Thus, for some D of size s' and every (Y, Z) with $\mathbf{H}_\infty(Y|Z) \geq k'$ we have

$$\mathbf{E}_{z \leftarrow Z} |\mathbf{E}_{x \leftarrow X|Z=z} D(x, z) - \mathbf{E}_{x \leftarrow Y|Z=z} D(x, z)| \geq \epsilon'. \quad (10)$$

Let D' be a distinguisher obtained from Lemma 7. Consider the following distinguisher D'' : on input (x, z) , which comes either from (X, Z) or (Y', Z) do the following:

- for $i = 1$ to $\ell = \lceil \frac{64}{\epsilon^2} \rceil - 1$ sample $y_i \leftarrow Y'|Z = z$ using the circuit Γ ,
- if $D'(x, z) > \max_{i=1, \dots, \ell} D'(y_i, z)$ — output 1, otherwise output 0.

Clearly D'' has complexity at most $(\ell + 1) \cdot (s' + \text{size}(\Gamma)) = s$. We will show that it gives sufficient distinguishing advantage. We start with the following easy observation, used implicitly already in [CKLR11] (the proof of Lemma 16).

Lemma 8. *For D be a $[0, 1]$ -valued function. If Y^+ is distributed uniformly over Max_D^k , then for any Y with $\mathbf{H}_\infty(Y) \geq k + \log \frac{1}{\delta}$ we have*

$$\mathbf{P}_{x \leftarrow Y} [D(x) - \mathbf{E}_{x \leftarrow Y^+} D(x) > 0] < \delta.$$

The proof that D'' is indeed a good distinguisher consists of two steps

Claim 6.1. *On input $(x, z) \leftarrow (X, Z)$ the circuit D'' outputs 1 with probability at least $\epsilon'^2/32$.*

Proof. Consider a distribution (Y^+, Z) such that for every z the distribution $Y^+|Z = z$ is uniform over $\text{Max}_{D(\cdot, z)}^k$. Since y_i are independent and distributed according to $Y'|Z = z$, it follows from Lemma 8 that $\mathbf{E}_{x \leftarrow Y^+|Z=z} D'(x) \geq \max_i D'(y_i, z)$ holds with probability at least

$$\left(1 - 2^{k'-k}\right)^\ell \geq 1 - \ell \cdot 2^{k'-k} \geq \frac{1}{2}.$$

Now, Lemma 7 yields $D'(x, z) > \mathbf{E}_{x \leftarrow Y^+|Z=z} D'(x)$ with probability at least $\frac{\epsilon'^2}{16}$ over (x, z) . Since sampling y_i is independent from (X, Z) , the claim follows. \square

Claim 6.2. On input $(y, z) \leftarrow (Y', Z)$ the circuit D'' outputs 1 with probability at most $\epsilon'^2/64$.

Proof. Note that y as well as the samples y_1, \dots, y_ℓ are all independent copies of the same distributions $Y'|Z = z$. Therefore probability that $y > \max_{i=1, \dots, \ell} y_i$ is at most $\frac{1}{\ell+1} \leq \frac{\epsilon'^2}{64}$. \square

From the last two claims we get the inequality $\mathbf{P}(D''(X, Z) = 1) - \mathbf{P}(D''(Y, Z) = 1) \geq \frac{1}{64}\epsilon'^2$, which completes the proof of Theorem 7. \square

6.3 Approximate counting

It turns out that using a technique called the *approximate counting*, one can show a conversion from metric to modulus entropy. However, we need some additional assumptions to achieve both: high accuracy and efficiency in the approximate counting:

Theorem 8. Suppose that one of the following is true:

- (a) $\mathbf{H}^{\text{Metric}, \text{rand}\{0,1\}, s, \epsilon}(X|Z) \geq k$ against circuits of size s ,
- (b) $\mathbf{H}^{\text{Metric}, \{0,1\}, s, \epsilon}(X|Z) \geq k$ against circuits of size $s = \text{poly}(n)$, with an access to an **NP**-oracle.

Then we have $\mathbf{H}^{\text{Metric}, s', \epsilon'}(X|Z) \geq k'$, where $\epsilon' = 8\sqrt{\epsilon}$, $k' = k - \log \frac{1}{\epsilon}$ and s' given by $s' = \mathcal{O}\left(s \cdot \frac{2^{k-n-2} \cdot \epsilon}{\log(1/\epsilon)}\right)$ in case (a) or $s' = \text{poly}(n, \epsilon)$ in case (b).

Note that to make the conversion in (a) efficient, we need the assumption that k is large as it is easy to see that if k is much smaller than n then, in the formula that gives the bound on s' , the 2^{k-n-2} factor starts to dominate over ϵ .

Proof of Theorem 8. Suppose that $\mathbf{H}^{\text{Metric}, s', \epsilon'}(X|Z) < k'$. Then Lemma 7 implies that for all $Y \in \{0, 1\}^n$ with $\mathbf{H}_\infty(Y|Z) \geq k'$ and some distinguisher D' of complexity $s' + 1$ we have

$$\mathbf{P}_{(x,z) \leftarrow (X,Z)} \left[D'(x, z) - \mathbf{E} D'(Y|Z = z, z) \geq \frac{\epsilon'}{4} \right] \geq \frac{\epsilon'^2}{16}. \quad (11)$$

Since

$$\max_{Y_z: \mathbf{H}_\infty(Y_z) \geq k'} \mathbf{E} D(Y_z, z) = \min \left(1, 2^{-k'} |D'(\cdot, z)| \right) \quad (12)$$

hence, combining it with (11), we obtain

$$\mathbf{P}_{(x,z) \leftarrow (X,Z)} \left[D'(x, z) - \frac{|D'(\cdot, z)|}{2^{k'}} \geq \frac{\epsilon'}{4} \right] \geq \frac{\epsilon'^2}{16}. \quad (13)$$

We now show that there exists a function h such that for every z the random variable $h(z)$ satisfies

$$\mathbf{P} \left(\left| h(z) - \frac{|D'(\cdot, z)|}{2^{k'}} \right| \leq \frac{\epsilon'}{8} \right) \geq 1 - \frac{\epsilon'^2}{64}. \quad (14)$$

and $h(z)$ is samplable for all z 's satisfying $|D'(\cdot, z)| < 2^{k'}$. More precisely: there there exists a randomized circuit of size $\mathcal{O}\left(s' \cdot \frac{2^{n-k}}{\epsilon^2} \log \frac{1}{\epsilon}\right) = s$, which computes $h(z)$ correctly for every such z . This is a corollary from following claim.

Claim 6.3. Let D be a boolean circuit such that $|D| \leq 2^k$. Then for $\delta', \delta'' \in (0, \frac{1}{2})$, $\ell > 4 \cdot 2^{n-k} \frac{1}{\delta'^2} \log \frac{1}{\delta''}$ and $(U_i)_{i=1, \dots, \ell}$ being independent and uniform, the following inequality holds:

$$\mathbf{P} \left[\left| \frac{1}{\ell} \sum_{i=1}^{\ell} D(U_i) - 2^{-n} |D| \right| \geq 2^{k-n} \delta' \right] \leq 2\delta''.$$

Proof of Claim 6.3. Define a random variable $g = \frac{1}{\ell} \sum_{i=1}^{\ell} D(U_i)$. The Chernoff Inequality⁵ yields

$$\mathbf{P}[|g - \mathbf{E}D(U)| \geq \delta] \leq 2 \max \left(e^{-\frac{\delta^2 \ell^2}{4\sigma^2}}, e^{-\frac{\ell\delta}{2}} \right),$$

where $\sigma^2 = \text{Var} \left(\sum_{i=1}^{\ell} D(U_i) \right)$. Since $\text{Var}(D(U_i)) = 2^{k-n}(1 - 2^{k-n})$ we have $\sigma^2 = \ell \cdot 2^{k-n}(1 - 2^{k-n})$.

By setting $2^{n-k}\delta = \delta'$ we get $\frac{\delta^2 \ell^2}{4\sigma^2} \geq \frac{2^{k-n}\ell\delta'^2}{4}$ and $\frac{\ell\delta}{2} \geq \frac{2^{k-n}\ell\delta'}{2}$. Choosing ℓ sufficiently large (so that $2^{k-n}\ell\delta'^2 > 4 \log \frac{1}{\delta'}$) we obtain $\mathbf{P}[|g \cdot 2^{n-k} - |D| \cdot 2^{-k}| \geq \delta'] \leq 2e^{-\log \frac{1}{\delta'}} < 2\delta'$. \square

Defining $h(z) = \frac{2^{k-n}}{\ell} \sum_{i=1}^{\ell} D(U_i, z)$, we obtain a required sampler for $h(z)$. Consider the following distinguisher D'' : on input (x, z) , which comes either from (X, Z) or (Y, Z) , return 1 iff $D'(x, z) > h(z) + \frac{\epsilon'}{8}$. We will prove that D'' distinguishes between (X, Z) and all (Y, Z) satisfying $\mathbf{H}_{\infty}(Y|Z) \geq k$. Note that if $D''(x, z) = 1$ then $h(z) < 1 - \frac{\epsilon'}{8}$ and hence $|D'(\cdot, z)| < 2^{k'}$. Especially, D'' is of complexity at most s . Now, inequalities (14) and (13) yield

$$\begin{aligned} \mathbf{P}_{(x,z) \leftarrow (X,Z)} \left[D'(x, z) > h(z) + \frac{\epsilon'}{8} \right] &\geq \\ \mathbf{P}_{(x,z) \leftarrow (X,Z)} \left[D'(x, z) > \frac{|D'(\cdot, z)|}{2^{k'}} + \frac{\epsilon'}{4} \right] - \frac{\epsilon'^2}{64} &\geq \frac{3}{4} \cdot \frac{\epsilon'^2}{16}, \end{aligned}$$

Choosing $k' = k + \log \frac{1}{\delta}$ where $\delta = \frac{\epsilon'^2}{64}$, using (12), (13), (14) and Lemma 8, we obtain

$$\begin{aligned} \mathbf{P}_{(x,z) \leftarrow (Y,Z)} \left[D'(x, z) > h(z) + \frac{\epsilon'}{8} \right] &\leq \\ \mathbf{P}_{(x,z) \leftarrow (Y,Z)} \left[D'(x, z) > \frac{|D'(\cdot, z)|}{2^{k'}} \right] + \frac{\epsilon'^2}{64} &\leq \frac{1}{2} \cdot \frac{\epsilon'^2}{16}. \end{aligned}$$

Combining the last two estimates yields, if only $\mathbf{H}_{\infty}(Y|Z) \geq k'$, the inequality

$$\mathbf{P}[D'(X, Z) = 1] - \mathbf{P}[D'(Y, Z) = 1] \geq \frac{\epsilon'^2}{64}$$

which completes the proof for case (a). In case (b), we proceed in the same way but we compute numbers $h(z)$ using an **NP** oracle. The basic result we use can be stated as follows:

Lemma 9. [OG09] *There is a probabilistic algorithm which, given a boolean circuit D over $\{0, 1\}^n$ of size $\text{poly}(n)$ and a natural number M , decides, with success probability at least $\frac{3}{4}$, whether $\frac{1}{4}M < |D| < 4M$, in time $\text{poly}(n)$, using an oracle for **NP**.*

Let us make three important observations:

- The success probability $\frac{3}{4}$ can be amplified to $1 - \delta$, by repeating the algorithm $\mathcal{O}(\log \frac{1}{\delta})$ times and taking the majority answer.
- The factor 4 can be improved to $1 + \gamma$, by running the algorithm on the circuit $D' = D_1 \wedge \dots \wedge D_k$, where D_i for $i = 1, \dots, k$ are copies of D and k is such that $(1 + \gamma)^k \leq 4$.

⁵We use the following version: Let X_i be a random variables satisfying $|X_i - \mathbf{E}X_i| \leq 1$ and $X = \sum_i X_i$. Then

$$\mathbf{P}[|X - \mathbf{E}X| \geq \lambda\sigma] \leq 2 \min \left(e^{-\frac{\lambda^2}{4}}, e^{-\frac{\lambda\sigma}{2}} \right), \text{ where } \sigma = \text{Var}(X)$$

Hence, there is an algorithm which, with probability at least $1 - \delta$, computes a value g such that $(1 - \gamma)M < |D| < (1 + \gamma)M$, in time $\text{poly}\left(n, \frac{1}{\gamma}, \log \frac{1}{\delta}\right)$, using an oracle for **NP**. For every z , let $M(z)$ be a value obtained by applying this algorithm to the circuit $D'(\cdot, z)$ and $\gamma = \frac{\epsilon'}{16}$, $\delta = 1 - \frac{\epsilon'^2}{64}$. Define $h(z) := 2^{-k}M(z)$. If $|D'(\cdot, z)| < 2^{k'}$, then $|M(z) - |D'(\cdot, z)|| \leq 2 \cdot 2^{k'} \cdot \frac{\epsilon'}{16}$ holds with probability at least $1 - \frac{\epsilon'^2}{64}$, and thus for such values z holds the same estimate as in (14). We proceed further with h as in the previous proof. \square

7 Some technical results

Lemma 10. *Let $X \in \{0, 1\}^n$ be a random variable, $f : \{0, 1\}^m \rightarrow \{0, 1\}^n$ be a deterministic function computable by a circuit of size s , and ϵ satisfy $0 < \epsilon < \frac{1}{12}$. Then*

$$\tilde{\mathbf{H}}^{\text{Metric}, \det\{0,1\}, s, \epsilon}(f(X)|X) < 3$$

Proof. Consider the following distinguisher D : on the input (y, x) , where $x \in \{0, 1\}^m$ and $y \in \{0, 1\}^n$, run $f(x)$ and return 1 iff $f(x) = y$. Then for every x we get $D(f(x), x) = 1$. Let Y be any random variable over $\{0, 1\}^n$ such that $\tilde{\mathbf{H}}_\infty(Y|X) \geq 3$. Then by Lemma 1 we obtain

$$\mathbf{H}_\infty(Y|X = x) \geq 3 - \log_2(3)$$

with probability $\frac{2}{3}$ over $x \leftarrow X$. Since $D(y, x) = 0$ if $y \neq x$, for any such x we have

$$\mathbf{E}_{y \leftarrow Y|X=x} D(y, x) \leq 2^{-(3 - \log_2(3))} \leq \frac{3}{8},$$

and thus, with probability $\frac{2}{3}$ over $x \leftarrow X$,

$$\overbrace{\mathbf{E}_{y \leftarrow f(X)|X=x} D(y, x)}^{=1} - \mathbf{E}_{y \leftarrow Y|X=x} D(y, x) \geq \frac{5}{8}$$

Taking the expectation over $x \leftarrow X$ we obtain finally

$$\mathbf{E}_{y, x \leftarrow f(X), X} D(y, x) - \mathbf{E}_{y, x \leftarrow Y, X} D(y, x) \geq \frac{2}{3} \cdot \frac{5}{8} - \frac{1}{3} \cdot 1 = \frac{1}{12}.$$

\square

We use the lemma above to show that the estimate in Lemma 3 cannot be improved:

Theorem 9 (Tightness of the estimate in Lemma 3). *Suppose that there exists an exponentially secure pseudorandom generator f . Then for every m and $C > 0$ we have*

$$\mathbf{H}^{\text{HILL}, \text{rand}\{0,1\}, 2^{\mathcal{O}(m)}, \frac{1}{2^{\mathcal{O}(m)}}}(f(U_m)) \geq m + C$$

and simultaneously,

$$\tilde{\mathbf{H}}^{\text{Metric}, \det\{0,1\}, \text{poly}(m), \frac{1}{\text{poly}(m)}}(f(U_m)|U_m) \leq 3$$

Proof. The first inequality follows directly from the definition of the exponentially secure pseudorandom generator. The second inequality is implied by Lemma 10. \square

Below we prove the equivalence between boolean and real valued distinguishers

Theorem 10. *For any random variables X, Z over $\{0, 1\}^n, \{0, 1\}^m$ we have*

$$\mathbf{H}^{\text{Metric}, \det[0,1], s', \epsilon}(X|Z) = \mathbf{H}^{\text{Metric}, \det\{0,1\}, s, \epsilon}(X|Z)$$

where $s' \approx s$.

Proof. We only need to prove $\mathbf{H}^{\text{Metric}, \det[0,1], s', \epsilon}(X|Z) \geq \mathbf{H}_{\infty}^{\text{Metric}, \det\{0,1\}, s, \epsilon}$ as the other direction is trivial (because the class $(\det[0,1], s)$ is larger than $(\det\{0,1\}, s)$). Suppose that $\mathbf{H}^{\text{Metric}, \det[0,1], s, \epsilon}(X|Z) < k$. Then for some D and all Y satisfying $\mathbf{H}_{\infty}(X|Z) \geq k$ we have

$$|\mathbf{E}_{(x,z) \leftarrow (X,Z)} D(x, z) - \mathbf{E}_{(x,z) \leftarrow (Y,Z)} D(x, z)| \geq \epsilon$$

Applying the same reasoning as in Thm. 6 we can replace D with D' which is equal either to D or D^c , obtaining, for all distributions $\mathbf{H}_{\infty}(Y|Z) \geq k$, the following:

$$\mathbf{E} D'(X, Z) - \mathbf{E} D'(Y, Z) \geq \epsilon.$$

Consider the distribution (Y^+, Z) minimizing the left side of the above inequality. Equivalently, it maximizes the expected value of D' under the condition $\mathbf{H}_{\infty}(Y|Z) \geq k$. Since this condition means that $\mathbf{H}_{\infty}(Y^+|Z = z) \geq k$ for all z , we conclude that $Y^+|Z = z$, for fixed z , is distributed over 2^k values of x giving the greatest values of $D'(x, z)$. Calculating the expected values in the last inequality via integration of the tail yields

$$\int_{t \in [0,1]} \mathbf{P}_{(x,z) \leftarrow (X,Z)} [D(x, z) > t] dt - \int_{t \in [0,1]} \mathbf{P}_{(x,z) \leftarrow (Y^+, Z)} [D(x, z) > t] dt \geq \epsilon$$

therefore for some number $t \in (0, 1)$, the following holds:

$$\mathbf{P}_{(x,z) \leftarrow (X,Z)} [D(x, z) > t] \geq \mathbf{P}_{(x,z) \leftarrow (Y^+, Z)} [D(x, z) > t] + \epsilon.$$

Let D'' be a $\{0, 1\}$ -distinguisher that for every (x, z) outputs 1 iff $D(x, z) > t$. Clearly D'' is of size $s + \mathcal{O}(1)$ and satisfies

$$\mathbf{E}_{(x,z) \leftarrow (X,Z)} D''(x, z) \geq \mathbf{E}_{(x,z) \leftarrow (Y^+, Z)} D''(x, z) + \epsilon.$$

We assumed that (Y, Z) maximizes $\mathbf{E} D'(Y, Z)$. Now we argue that (Y, Z) is also maximal for D'' . We know that for every z the distribution Y_z is flat over the set $\text{Max}_{D'(\cdot, z)}^k$ of 2^k values of x corresponding to largest values of $D'(x, z)$. It is easy to see that $\text{Max}_{D'(\cdot, z)}^k = \text{Max}_{D''(\cdot, z)}^k$. Therefore, we have shown in fact that

$$\mathbf{E}_{(x,z) \leftarrow (X,Z)} D''(x, z) - \max_{(Y,Z): \mathbf{H}_{\infty}(Y|Z) \geq k} \mathbf{E}_{(x,z) \leftarrow (Y,Z)} D''(x, z) \geq \epsilon,$$

which means exactly that $\mathbf{H}^{\text{Metric}, \{0,1\}, s', \epsilon}(X|Z) < k$. □

References

- [BSW03] Boaz Barak, Ronen Shaltiel, and Avi Wigderson, *Computational analogues of entropy*, RANDOM-APPROX (Sanjeev Arora, Klaus Jansen, José D. P. Rolim, and Amit Sahai, eds.), Lecture Notes in Computer Science, vol. 2764, Springer, 2003, pp. 200–215.
- [CKLR11] Kai-Min Chung, Yael Tauman Kalai, Feng-Hao Liu, and Ran Raz, *Memory delegation*, Cryptology ePrint Archive, Report 2011/273, 2011, <http://eprint.iacr.org/>.
- [DORS08] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith, *Fuzzy extractors: How to generate strong keys from biometrics and other noisy data*, SIAM J. Comput. **38** (2008), no. 1, 97–139.
- [DP08] Stefan Dziembowski and Krzysztof Pietrzak, *Leakage-resilient cryptography in the standard model*, IACR Cryptology ePrint Archive **2008** (2008), 240.

- [FOR12] Benjamin Fuller, Adam O'Neill, and Leonid Reyzin, *A unified approach to deterministic encryption: New constructions and a connection to computational entropy*, Cryptology ePrint Archive, Report 2012/005, 2012, <http://eprint.iacr.org/>.
- [FR11] Benjamin Fuller and Leonid Reyzin, *Computational entropy and information leakage*, 2011, Master Thesis.
- [GW10] Craig Gentry and Daniel Wichs, *Separating succinct non-interactive arguments from all falsifiable assumptions*, Cryptology ePrint Archive, Report 2010/610, 2010, <http://eprint.iacr.org/>.
- [HILL99] Johan Hastad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby, *A pseudo-random generator from any one-way function*, SIAM J. Comput. **28** (1999), no. 4, 1364–1396.
- [OG09] Ryan O'Donnell and Venkatesan Guruswami, *An intensive introduction to computational complexity theory*, University Lecture, 2009, <http://www.cs.cmu.edu/~odonnell/complexity/>.
- [Rey11] Leonid Reyzin, *Some notions of entropy for cryptography*, Information Theoretic Security (Serge Fehr, ed.), Lecture Notes in Computer Science, vol. 6673, Springer Berlin Heidelberg, 2011, pp. 138–142.
- [Sha48] C. E. Shannon, *A mathematical theory of communication*, Bell system technical journal **27** (1948).
- [SK12] Akshay Wadia Stephan Krenn, Krzysztof Pietrzak, *A counterexample for the chain rule for conditional hill entropy and what deniable encryption has to do with it*, Manuscript, 2012.
- [VZ12] Salil Vadhan and Colin Jia Zheng, *Characterizing pseudoentropy and simplifying pseudorandom generator constructions*, Proceedings of the 44th symposium on Theory of Computing (New York, NY, USA), STOC '12, ACM, 2012, pp. 817–836.
- [Yao82] Andrew C. Yao, *Theory and application of trapdoor functions*, Proceedings of the 23rd Annual Symposium on Foundations of Computer Science (Washington, DC, USA), SFCS '82, IEEE Computer Society, 1982, pp. 80–91.